(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平8－63187

(43)公開日　平成８年(1996)３月８日

(51)Int.Cl.[6]		識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ１０Ｌ	5/02	Ｊ			
	3/00	Ｈ			

審査請求　未請求　請求項の数11　ＯＬ　（全９頁）

(21)出願番号　特願平6－195178

(22)出願日　平成６年(1994)８月19日

(71)出願人　000005223
富士通株式会社
神奈川県川崎市中原区上小田中1015番地

(72)発明者　片江　伸之
神奈川県川崎市中原区上小田中1015番地
富士通株式会社内

(74)代理人　弁理士　井桁　貞一

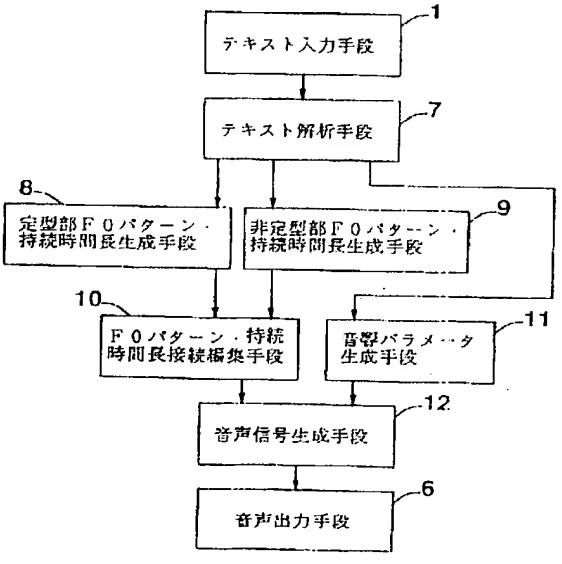
(54)【発明の名称】　音声合成装置

(57)【要約】

【目的】　本発明は、音声合成装置、特に交通情報や天気概況の音声サービスなどに用いる、定型文音声を合成するための音声合成装置に関し、聞き取りやすく、自然な韻律をもつ音声を合成することを目的とする。

【構成】　合成すべき一群のメッセージのすべてに共通する固定情報である定型部と該一群のメッセージ毎に異なる可変の情報である非定型部からなる文を音節、音素等の合成単位を滑らかにつなぎ合わせて合成する音声合成装置において、有声音の音声が包含する最低周波数である基本周波数の時間変化パターンであるＦ０パターンの生成にあたって、定型部のＦ０パターンを生成する第１のＦ０パターン生成手段と、非定型部のＦ０パターンを生成する第２のＦ０パターン生成手段と、当該各生成手段により生成したＦ０パターンを順次接続して文のＦ０パターンを生成する手段と、該Ｆ０パターンを用いて音声信号を合成する手段とを備えることを特徴とする音声合成装置を構成する。

本発明の原理図



1

【特許請求の範囲】

【請求項１】　合成すべき一群のメッセージに共通する固定情報と該一群のメッセージ毎に異なる可変情報をつなぎ合わせて一群のメッセージを合成する音声合成装置において、

基本周波数の時間変化パターンの生成にあたって、固定情報の基本周波数の時間変化パターンを生成する第１の生成手段と、可変情報の基本周波数の時間変化パターンを生成する第２の生成手段と、当該各生成手段により生成した基本周波数の時間変化パターンを順次接続して文の基本周波数の時間変化パターンを生成する編集手段と、該編集手段で生成された基本周波数の時間変化パターンを用いて音声信号を合成する合成手段とを備えることを特徴とする音声合成装置。

【請求項２】　請求項１に記載の第１の生成手段は、自然音声より抽出した固定情報の基本周波数の時間変化パターンを、基本周波数の時系列の形式を用いて記憶する手段と、入力文に適切な基本周波数の時系列を該記憶手段より読み込む手段とを備えることにより、基本周波数の時間変化パターンを生成することを特徴とする音声合成装置。

【請求項３】　請求項１に記載の第１の生成手段は、自然音声より抽出した固定情報の基本周波数の時間変化パターンを、該基本周波数の時間変化パターンを近似したモデルのパラメータの形式を用いて記憶する手段と、入力文に適切なパラメータを該憶記する手段より読み込む手段と、該パラメータより基本周波数の時系列を生成する手段を備えることにより、基本周波数の時間変化パターンを生成することを特徴とする音声合成装置。

【請求項４】　請求項１に記載の第２の生成手段は、可変情報の音節数とアクセント型の組合せについて自然音声より抽出した基本周波数の時間変化パターンを、基本周波数の時系列の形式を用いて記憶する手段と、入力文に適切な基本周波数の時系列を該記憶手段より選択し読み込む手段とを備えることにより、基本周波数の時間変化パターンを生成することを特徴とする音声合成装置。

【請求項５】　請求項１に記載の第２の生成手段は、可変情報の音節数とアクセント型のすべての組合せについて自然音声より抽出した基本周波数の時間変化パターンを、該基本周波数の時間変化パターンを近似したモデルのパラメータの形式を用いて記憶する手段と、入力に適切なパラメータを該記憶手段より選択し読み込む手段と、該パラメータより基本周波数の時系列を生成する手段を備えることにより、基本周波数の時間変化パターンを生成することを特徴とする音声合成装置。

【請求項６】　請求項１に記載の第２の生成手段は、可変情報の基本周波数の時間変化パターンを規則によって生成する手段を持つことを特徴とする音声合成装置。

【請求項７】　合成単位の各時間長の系列である持続時間長の生成にあたって、

2

固定情報の持続時間長を生成する第１の生成手段と、可変情報の持続時間長を生成する第２の生成手段と、当該各生成手段により生成した持続時間長を順次接続して、文の持続時間長を生成する編集手段と、該持続時間長を用いて音声信号を合成する手段とを備えることを特徴とする音声合成装置。

【請求項８】　請求項７に記載の第１の生成手段は、自然音声より抽出した固定情報の持続時間長を記憶する手段と、入力文に適切な持続時間長を該記憶手段より読み込む手段とを備えることによって、持続時間長を生成することを特徴とする音声合成装置。

【請求項９】　請求項７に記載の第２の生成手段において、

可変情報の持続時間長を生成する生成手段を持つことを特徴とする音声合成装置。

【請求項１０】　請求項１または請求項７に記載の音声合成装置において、当該音声合成装置が固定情報を提示し、ユーザが可変情報の入力および編集を行なうユーザインターフェイスを用いて合成文を入力することによって、固定情報と可変情報を分離することを可能にするテキスト入力手段を備えることを特徴とする音声合成装置。

【請求項１１】　請求項１または請求項７に記載の音声合成装置において、

当該音声合成装置が固定情報の提示と可変情報の入力候補の提示を行ない、該候補の可変情報を指定する選択手段と、固定情報と可変情報を分離することを可能にするテキスト入力手段を備えることを特徴とする音声合成装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、音声合成装置に関し、特に交通情報や天気概況の音声サービスなどに用いる、合成すべき一群のメッセージのすべてに共通する固定情報（以下、定型部と呼ぶ。）とメッセージ群で共通しない可変情報（以下、非定型部と呼ぶ。）からなる音声を合成する音声合成装置に関する。

【０００２】近年、社会一般の省力化・機械化の要請が益々強くなり、各種音声サービスの分野も例外ではなく、現在、交通情報や天気概況の音声サービス、銀行の振り込み照会サービスなどに、音声合成装置が使用されている。このため、音声合成装置は聞取りやすく、自然な韻律をもつ合成音声を提供する必要がある。

【０００３】

【従来の技術】従来の音声合成装置では、定型部には、あらかじめ録音しておいた音声を再生する録音編集方式、あるいは該音声をなんらかの音声パラメータに変換したものを蓄積しておき、そのパラメータから音声を合成する分析合成方式が用いられている。また、固有名詞や数字などの非定型部は、文字列から規則を用いて、音

声を生成する規則合成方式を用い、それぞれの方式で合成した音声を接続して、あるいは切替えて出力するのが一般的であった。

【０００４】従来技術による音声合成装置の構成図を図９に示す。　図中、１はテキスト入力手段、２はテキスト解析手段、３は定型部合成手段、４は非定型部合成手段、５は出力音声接続手段、６は音声出力手段をそれぞれ示す。　テキスト入力手段１に入力されたテキストを、テキスト解析手段２において、単語辞書を参照しながら解析する。その結果、定型部の部分は定型部合成手段３に入力され、蓄積してある定型部音声データから音声を合成する。可変な情報からなる部分は、非定型部合成手段４に入力され、文字列からの規則合成を行なう。それぞれの合成手段で合成した音声を、文として継るように出力音声接続手段５で接続し、音声出力手段６を介して出力する。

【０００５】

【発明が解決しようとする課題】ところが、音声の品質を見ると、規則合成方式の音声品質は録音編集方式や分析合成方式に比べて劣っているのが現状である。

【０００６】従って、録音編集方式または分析合成方式による定型部と規則合成方式による非定型部とを接続した音声では、定型部と非定型部の品質にギャップがあり、文中の重要な情報を含む非定型部が聞き取りにくいという問題があった。これに対して、文全体を同じ品質で生成するほうが聞き取りやすく、特に近年、技術の改良によって規則合成方式の音声品質が向上してきたこともあり、すべてを規則合成によって合成しても、十分に実用に耐えうるようになってきた。もちろん、すべて規則合成方式を用いれば、定型部を変更したい場合でも、音声を再収録する手間も省ける。

【０００７】ところで、我々が日常生活に用いている漢字かな混じり文から音声を合成するとき、規則合成方式では録音編集方式や分析合成方式とは異なり、辞書と規則を参照しながら、自然な韻律（イントネーション、アクセント、ポーズ等）を生成する必要がある。この過程で以下の２個の問題が存在する。

【０００８】第１の問題は、漢字かな混じり文を解析して表音文字列を生成する過程におけるものである。ここで、表音文字列とは、音素（日本語ではローマ字表記とほぼ等しい。）列または音節（日本語では仮名文字表記とほぼ等しい。）列に、ポーズ位置、アクセントの位置を示す表記を含めた文字列のことである。日本語は単語でわかち書きされておらず、漢字には幾通りもの読み方があるため、辞書と規則から表音文字列を生成しようとすると、誤読やアクセントの誤り、不自然なポーズの挿入などが頻繁に起こる。　第１の問題は、韻律情報を含む予め作成した入力文字列を記憶した記憶手段としての音声変換用入力列ファイルから抽出した文字列規則合成することにより解決されている（特開平4-107598参照。）が、構成費用の低減が要求され。

【０００９】第２の問題は、表音文字列から音響的（物理的）なパラメータを生成する過程におけるものである。例えば、イントネーションは声の高さの変化であり、有声音の音声が包含する最低周波数である基本周波数の時間変化パターン（以下、ＦＯパターンと称する。）を用いて制御するのが一般的である。これは数ミリ秒(msec)毎の基本周波数の時系列で表される。上記の表音文字列からこのＦＯパターンを生成するための規則として、有名なものに、藤崎モデルや点ピッチモデルなどがあるが、人間の複雑な発声機構や、内容、意味によっても微妙に変化するＦＯパターンを簡単な規則によって求めるのは困難である。また、発声がつかえたり間延びしたりせずに自然になるように、各音素あるいは音節の時間長を適切な値に設定している。ところが、この時間長は音素あるいは音節の種類によって一意に決まるものではなく、この音素あるいは音節が置かれている文中の位置や周辺の音韻環境によって複雑に影響されるものであり、これもまた単純な規則では求まらないものである。

【００１０】

【課題を解決するための手段】図２は本発明の概念図である。　以下、同図と「今夜の［東京］地方の天気は［晴れ］でしょう。」という例文によって説明する。

【００１１】本文は「今夜の・・地方の天気は・・・でしょう。」という定型部と「東京」「晴れ」という非定型部から構成されており、非定型部はそれぞれ「神奈川県」「雨」のような単語と置換することが可能であるとする。このような文を合成するときに、定型部に関しては、同文を人間が発声した音声から定型部のＦＯパターンや持続時間長を抽出し、例えば、ＦＯパターンであれば数msec毎の基本周波数値の時系列として、持続時間長であれば各音素の長さの系列として蓄積しておく。非定型部に関しては、非定型部への入力が期待される単語あるいは文節などの音節数とアクセント型のすべての組合せのＦＯパターンを蓄積しておき、入力文、またはそれを解析した表音文字列から、同じ音節数とアクセント型の組合せのＦＯパターンを読み込む。このＦＯパターンは、音節数とアクセント型だけでなく、文全体のＦＯパターンの中で決まるものであるから、定型部のいずれの位置に挿入するかによって、ＦＯパターンはそれぞれ異なるものを持ち、選択する必要がある。たとえば、「東京」という単語であれば４モーラ０型であるから、定型部の「今夜の・・・地方」の位置に挿入されるパターンの中から４モーラ０型のＦＯパターンを選択する。非定型部の持続時間長は規則により生成する。定型部と非定型部に分けて検索した（あるいは生成した）ＦＯパターンと持続時間長を順に接続することによって、文全体のＦＯパターンを作成する。ＦＯパターンは、文全体で連続して接続される。

5

【００１２】また、非定型部に関してＦ０パターンを蓄積しておかずに、規則によって生成しても、文全体のＦ０パターンをすべて規則で生成した場合よりも高品質な音声が得られる。

【００１３】

【作用】本発明の原理図を図１に示す。　図中、１はテキスト入力手段、７はテキスト解析手段、８は定型部Ｆ０パターン・持続時間長生成手段、９は非定型部Ｆ０パターン・持続時間長生成手段、１０はＦ０パターン・持続時間長接続編集手段、１１は音響パラメータ生成手段、１２は音声信号生成手段、６は音声出力手段をそれぞれ示す。　テキスト入力手段１に合成するテキストが入力される。テキスト解析手段７では、入力テキストを非定型部と定型部に分離する。入力されたテキストが通常の漢字かな混じり文の場合は、定型部と非定型部に分離するために、任意文の規則合成に用いるようなテキスト解析が必要であるが、ユーザインタフェースによって、定型部と非定型部を分けて入力できる場合には、単純に定型部と非定型部をそれぞれのＦ０パターン・持続時間長生成手段に出力するだけでよい。又、テキスト解析手段７では入力文から表音文字列（音素列または音節列）を生成して音響パラメータ生成手段１１に出力する。定型部については定型部Ｆ０パターン・持続時間長生成手段８において、非定型部については非定型部Ｆ０パターン・持続時間長生成手段９において、それぞれ、Ｆ０パターンおよび持続時間長を生成する。これらのＦ０パターンおよび持続時間長は、Ｆ０パターン・持続時間長接続編集手段１０において順次接続され、文全体のＦ０パターンおよび持続時間長が生成される。音響パラメータ生成手段１１では、音素列または音節列などの表音文字列を基に、ホルマント等の音響パラメータを生成する。音響パラメータは音声信号生成手段１２に用いる合成方式によって決まる。また、合成方式としては波形を直接編集する波形編集方式があり、この方式を用いた場合は音響パラメータではなく、それに相当するものとして、波形接続情報を生成することになるが、ここでは、音響パラメータに含めて扱う。音声信号生成手段１２では、Ｆ０パターン、持続時間長、および音響パラメータから、音声信号を生成し、音声出力手段６から出力する。

【００１４】

【実施例】Ｆ０パターン生成方法には３つのレベルが考えられる。第１のレベルは、自然音声から抽出したＦ０パターンをそのまま基本周波数の時系列の形式で蓄積しておき合成時に読み込む方法であり、最も自然な音声の合成が期待されるものである。第２のレベルは、自然音声のＦ０パターンをモデルにより近似して、そのモデルのパラメータを蓄積しておき、合成時にパラメータから基本周波数の時系列の形式に変換する方法である。第３のレベルは、テキスト解析結果からモデルのパラメータを規則的に生成し、該パラメータから基本周波数の時系列を生成する方法である。

6

【００１５】また、持続時間長生成方法には２つのレベルが考えられる。第１のレベルは、自然音声から抽出した持続時間長をそのまま時間長の系列として蓄積しておき合成時に読み込む方法である。第２のレベルは、上記の時間長をテキスト解析結果から規則的に生成する方法である。非定型部と定型部のＦ０パターンおよび持続時間長生成方法として、上記のレベルそれぞれの組合せが考えられる。これらを実施例として以下に述べる。

【００１６】本発明の第１の実施例の構成図を図３に示す。　本実施例は特許の請求項２、４、８および９に対応している。図中、０１１はテキスト入力部、７１はテキスト解析部、７２は定型／非定型判定部、７３は出力切替部、７４は単語辞書、７５は定型部文例蓄積部、８１は定型部持続時間長読み込み部、８２は定型部Ｆ０パターン読み込み部、８３は定型部持続時間長蓄積部、８４は定型部Ｆ０パターン蓄積部、９１は非定型部持続時間長生成部、９２は非定型部Ｆ０パターン読み込み部、９３はアクセント辞書、９４は非定型部Ｆ０パターン蓄積部、１０１は持続時間長接続編集部、１０２はＦ０パターン接続編集部、１１１は音響パラメータ生成部、１１２は音響パラメータ蓄積部、１２１は音声信号生成部、６１は音声出力部を示す。

【００１７】あらかじめ、定型部について自然音声より抽出した定型部Ｆ０パターンを定型部Ｆ０パターン蓄積部８４に格納し、非定型部について、その音節数とアクセント型のすべての組合せの非定型部Ｆ０パターンを非定型部Ｆ０パターン蓄積部９４に格納し、定型部について自然音声より抽出した定型部持続時間長を定型部持続時間長蓄積部８３に格納してある。合成するテキストがテキスト入力部０１１に入力される。入力が漢字かな混じり表記である場合は、テキスト解析部７１において、単語辞書７４を参照しながら、テキストを解析する。定型／非定型判定部７２では、定型部文例蓄積部７５に格納されている定型文例を参照し、解析結果を定型部と非定型部に分離する。出力切替部７３は定型部と非定型部をそれぞれの持続時間長、Ｆ０パターン生成部に出力する。またこのとき、テキストを解析した結果として、入力テキストの表音文字列（音素列または音節列など）を音響パラメータ生成部１１１に出力する。

【００１８】定型部については、定型部持続時間長読み込み部８１において、定型部持続時間長蓄積部８３から持続時間長を読み込み、又、定型部Ｆ０パターン読み込み部８２において、定型部Ｆ０パターン蓄積部８４からＦ０パターンを読み込み、それぞれ持続時間長接続編集部１０１を経由し、Ｆ０パターン接続編集部１０２に出力する。非定型部については、非定型部持続時間長生成部９１において、規則により持続時間長を生成する。規則による持続時間長生成は、非定型部の各音素または音

節について時間長テーブルを検索し、音素環境などによって補正するといった方法がとられるのが一般的である。次に、非定型部Ｆ０パターン読み込み部９２では、非定型部の単語のアクセントをアクセント辞書９３から獲得し、音節数とアクセント型から非定型部Ｆ０パターン蓄積部９４を参照して、読み込んだＦ０パターンを持続時間長接続編集部１０１、Ｆ０パターン接続編集部１０２に出力する。持続時間長接続編集部１０１では、定型部と非定型部それぞれの音素時間長を順番に接続し、文全体の持続時間長の系列を作成する。Ｆ０パターン接続編集部１０２では、定型部と非定型部のそれぞれのＦ０パターンを順番に接続し、文全体のＦ０パターンを作成する。Ｆ０パターンは発声中連続であるので、二つの定型部と非定型部で読み込んだＦ０パターンのそれぞれに不連続がある場合には、適切なスムージングを行なうなどの編集を行なわなければならない。

【００１９】一方、音響パラメータ生成部１１１では、入力の表音文字列をもとに音響パラメータを生成する。音響パラメータ蓄積部１１２には、音響パラメータが格納されている。ここで言う、音響パラメータとは、データ容量を圧縮するために音声生成モデルを用いて音声データを数値化したものであり、ホルマント、ＰＡＲＣＯＲ、ＬＳＰなどの種類があり。これらの音響パラメータを用いた合成方式を、それぞれホルマント合成、ＰＡＲＣＯＲ合成、ＬＳＰ合成と呼び、音声信号生成部１２１によって実現される。　また、合成方式としては波形を直接編集する波形編集方式があり、この方式を用いた場合は音響パラメータではなく、それに相当するものとして、波形接続情報を生成することになるが、ここでは、音響パラメータに含めて扱う。音響パラメータは、表音文字ごと、あるいはそれを前後の音素環境などにより細分化した単位で蓄積されている。表音文字列にしたがってこれを読み込み、連接することによって、合成文の音響パラメータ列が生成される。音声信号生成部１２１では、以上で生成された合成文の持続時間長、Ｆ０パターン、音響パラメータ列より音声信号を生成する。音声出力部６１では、その音声信号をＤＡ変換することにより、合成音声として出力する。

【００２０】本発明の第２の実施例の構成図を図４に示す。　本実施例は特許の請求項３および５に対応している。本実施例は、実施例１の定型部Ｆ０パターン読み込み部８２と定型部Ｆ０パターン蓄積部８４を定型部Ｆ０パラメータ読み込み部８５、定型部Ｆ０パターン生成部８６、および定型部Ｆ０パラメータ蓄積部８７に、また、非定型部Ｆ０パターン読み込み部９２と非定型部Ｆ０パターン蓄積部９４を非定型部Ｆ０パラメータ読み込み部９５、非定型部Ｆ０パターン生成部９６、および非定型部Ｆ０パラメータ蓄積部９７に置き換えたものである。

【００２１】本実施例では、あらかじめ、自然音声から

抽出したＦ０パターンをモデルにより近似して、そのパラメータを定型部Ｆ０パラメータ蓄積部８７と非定型部Ｆ０パラメータ蓄積部９７に蓄積しておく。音声を合成する際に、定型部に関しては、定型部Ｆ０パラメータ読み込み部８５において、定型部のＦ０パラメータを定型部Ｆ０パラメータ蓄積部８７から読みだし、定型部Ｆ０パターン生成部８６において、パラメータから基本周波数の時系列（Ｆ０パターン）を生成する。同様に、非定型部についても、非定型部Ｆ０パラメータ読み込み部９５において、非定型部の単語のアクセントをアクセント辞書９３から獲得し、その音節数とアクセント型によって、非定型部Ｆ０パラメータ蓄積部９７から適切なＦ０パラメータを読みだし、非定型部Ｆ０パターン生成部９６において、パラメータから基本周波数の時系列（Ｆ０パターン）を生成する。

【００２２】本発明の第３の実施例の構成図を図５に示す。本実施例は特許の請求項６に対応している。本実施例は、実施例１の非定型部Ｆ０パターン読み込み部９２と非定型部Ｆ０パターン蓄積部９４を非定型部Ｆ０パターン生成部９８に置き換えたものである。その他の部分の処理は実施例１と同様であるから、非定型部Ｆ０パターン生成部９８についてのみ説明する。

【００２３】非定型部Ｆ０パターン生成部９８では、非定型部の単語のアクセントをアクセント辞書９３から獲得し、文中の位置などを考慮してＦ０パターンを規則により生成する。Ｆ０パターンを規則により生成する方法としては、藤崎モデルや点ピッチモデルなどのモデルを用いる方式が一般的であり、この場合もこれらが応用できる。

【００２４】本発明の第４の実施例の構成図を図６に示す。　本実施例は請求項１０および１１に対応している。本実施例は、実施例１のテキスト入力部のユーザインタフェイスを置き換えることで、テキストの解析をより正確にしたものである。入力インターフェイス部０１２では、定型部文例蓄積部０１３より定型部を読みだし、ユーザインタフェースとして、図７または図８のように表示する。図７では、定型部には表示のみの機能しかないカラムを、非定型部には、自由に単語の入力／編集ができるエディット機能のあるカラムを用意し、使用者に非定型部の入力を促す。このようなインターフェイスで入力すると、定型部と非定型部の判定が不必要で、定型部のみを単語辞書７４で検索することによって、テキスト解析が可能である。

【００２５】図８では定型文例蓄積部１３に、非定型部の入力候補を蓄積しておき、非定型部のカラムを指定するとその箇所に入るべき入力候補が表示され、候補選択手段を用いて、いずれを入力とするか指定できるというインターフェイスを持っている。こちらも同様に、定型部と非定型部の判定が不必要で、定型部のみを単語辞書７４で検索することによって、テキスト解析が可能であ

る。以降の処理は他の実施例と同様である。

【００２６】

【発明の効果】以上説明した様に、本発明によれば、交通情報や天気概況の音声サービスなどに用いる、定型文音声を合成するための音声合成装置において、聞き取りやすく、自然な韻律をもつ音声を合成することができる。

【図面の簡単な説明】

【図１】　本発明の原理図である。

【図２】　本発明の基本的な考え方を示した概念図である。

【図３】　本発明の第１の実施例である。

【図４】　本発明の第２の実施例である。

【図５】　本発明の第３の実施例である。

【図６】　本発明の第４の実施例である。

【図７】　本発明のユーザインターフェースの第１の例である。

【図８】　本発明のユーザインターフェースの第２の例である。

【図９】　従来例である。

【符号の説明】

１　テキスト入力手段

２、７　テキスト解析手段

３　定型部合成手段

４　非定型部合成手段

５　出力音声接続手段

６　音声出力手段

８　定型部Ｆ０パターン・持続時間長生成手段

９　非定型部Ｆ０パターン・持続時間長生成手段

１０　Ｆ０パターン・持続時間長接続編集手段（編集手＊段と略す。）

１１　音響パラメータ生成手段

１２　音声信号生成手段

６１　音声出力部

７１、７１'　テキスト解析部

７２　定型／非定型判定部

７３　出力切替部

７４　単語辞書

７５、０１３　定型部文例蓄積部

８１　定型部持続時間長読み込み部

８２　定型部Ｆ０パターン読み込み部

８３　定型部持続時間長蓄積部

８４、定型部Ｆ０パターン蓄積部

８５　定型部Ｆ０パラメータ読み込み部

８６　定型部Ｆ０パターン生成部

８７　定型部Ｆ０パラメータ蓄積部

９１　非定型部持続時間長生成部

９２　非定型部Ｆ０パターン読み込み部

９３　アクセント辞書

９４　非定型部Ｆ０パターン蓄積部

９５　非定型部Ｆ０パラメータ読み込み部

９６、９８　非定型部Ｆ０パターン生成部

９７　非定型部Ｆ０パラメータ蓄積部

０１１　テキスト入力部

０１２　入力インターフェース部

１０１　持続時間長接続編集部

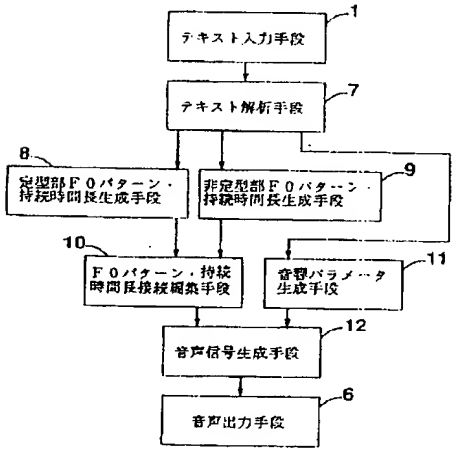
１０２　Ｆ０パターン接続編集部

１１１　音響パラメータ生成部

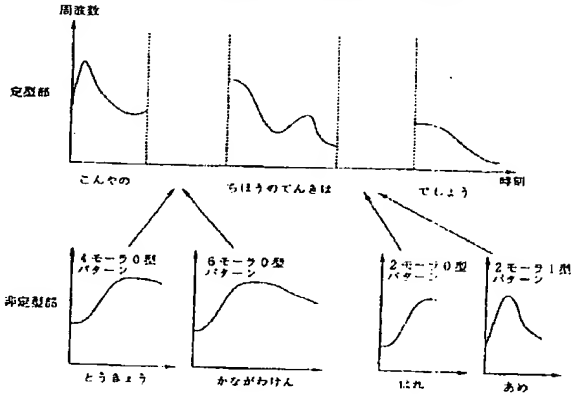
１１２　音響パラメータ蓄積部

１２１　音声信号生成部

【図１】

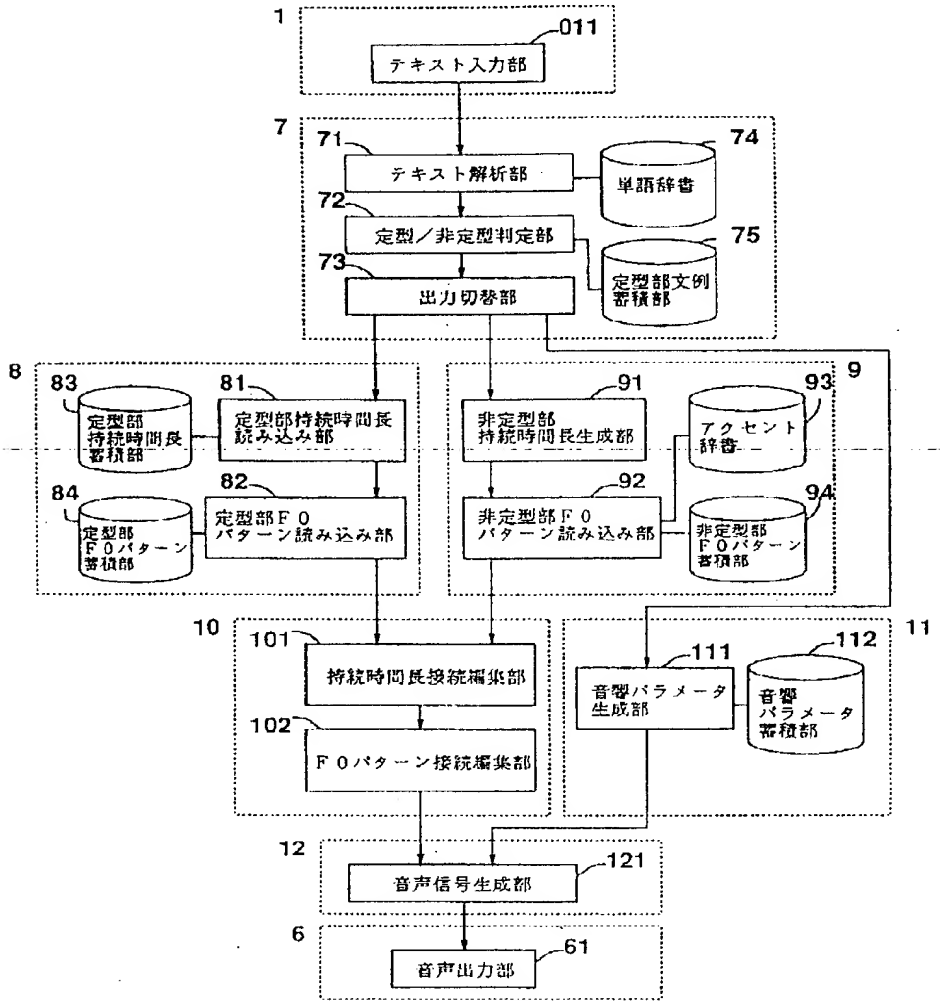


【図２】



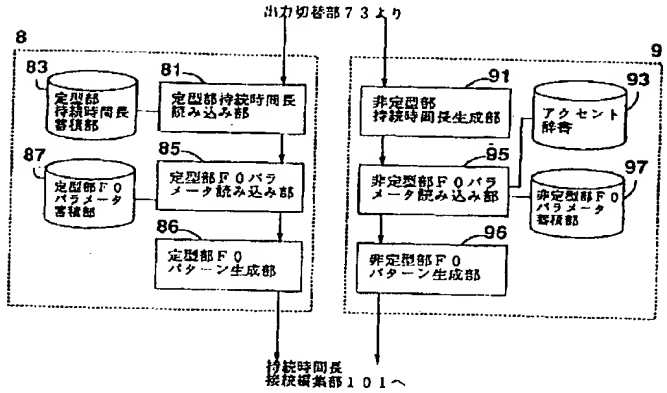
【図３】

本発明の第１の実施例



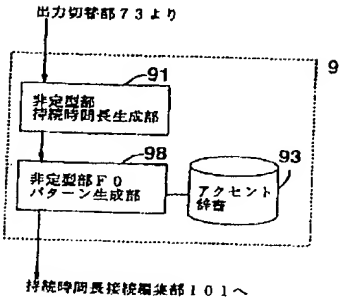
【図４】

本発明の第２の実施例



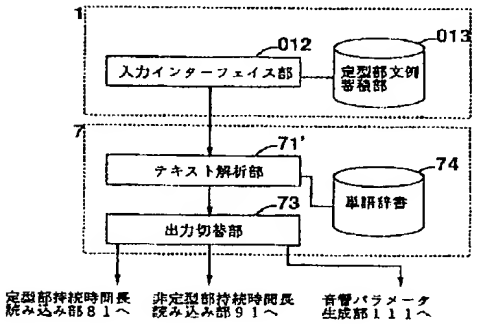
【図５】

本発明の第３の実施例



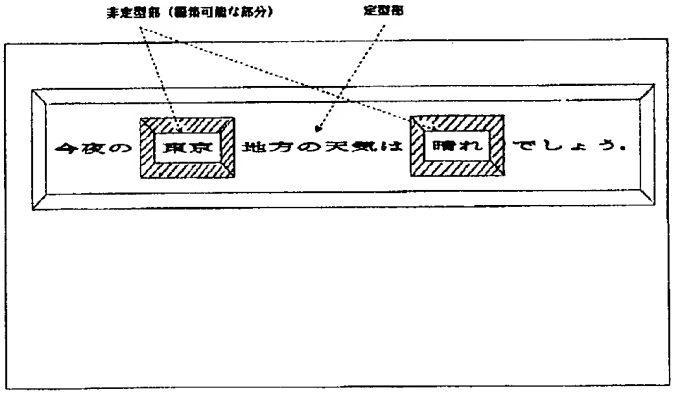
【図６】

本発明の第４の実施例



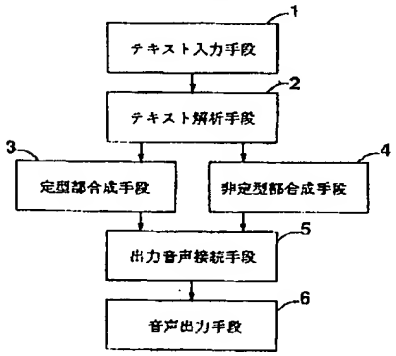
【図７】

本発明のユーザインターフェースの第１の例



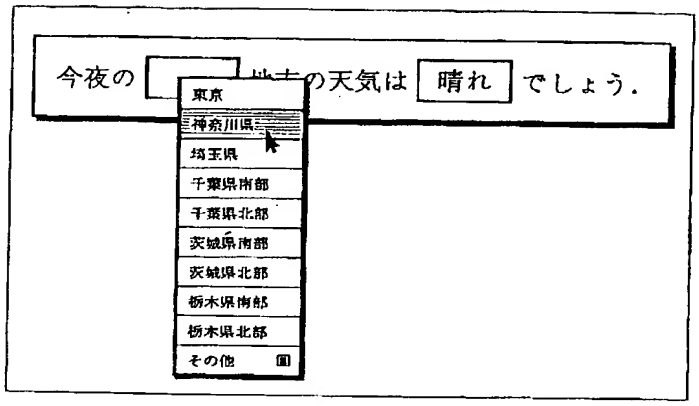
【図９】

従来例



【図８】

本発明のユーザインターフェースの第２の例



19)Japanese Patent Office

11)JP Unexamined Patent Publication H8-63187

43)March 8, 1996

54)[Title of the Invention] Voice Synthesizing Device

57)[Abstract]

[Purpose]

The present invention relates to a voice synthesizing device, and particularly to a voice synthesizing device for synthesizing voices of fixed sentences used for voice services such as traffic information or general weather condition, and it is an object thereof to synthesize voices with natural meters that are easy to hear.

[Structure]

In a voice synthesizing device for synthesizing sentences comprised of fixed form parts representing fixed information that are common to all of a group of messages to be synthesized and non-fixed form parts representing variable information that differ for each of the group of messages by smoothly combining synthesizing units such as syllables or phonemes, the voice synthesizing device is characterized by comprising, for generating F0 patterns which are time-varying patterns of basic frequencies having minimum frequencies included in voices of voiced sounds, a first F0 pattern generating means of generating F0 patterns for the fixed form parts, a second F0 pattern for generating F0 patterns for the non-fixed parts, a means of

generating a F0 pattern for a sentence upon sequentially connecting the F0 patterns that have been generated by the respective generating means, and a means of synthesizing voice signals by using the F0 pattern.

[What is claimed is]

[Claim 1]

A voice synthesizing device for synthesizing a group of messages upon combining fixed information that are common to the group of messages to be synthesized and variable information that differ for each of the group of messages, the voice synthesizing device comprising:

a first generating means of generating time-varying patterns of basic frequencies for the fixed information, a second generating means of generating time-varying patterns of basic frequencies for the variable information, an editing means of generating a time-varying pattern of basic frequency for a sentence upon sequentially connecting the time-varying patterns of basic frequencies generated by the respective generating means, and a synthesizing means of synthesizing a voice signal by using the time-varying pattern of basic frequency generated by the editing means.

[Claim 2]

A voice synthesizing device, wherein the first generating means according to Claim 1 generates the time-varying patterns of basic frequencies by comprising a means of storing the

time-varying patterns of basic frequencies for the fixed information extracted from natural voices upon employing a style of chronological orders of basic frequencies and a means of reading a chronological order of basic frequency that suits a sentence to be inputted from the storing means.

[Claim 3]

A voice synthesizing device, wherein the first generating means according to Claim 1 generates the time-varying patterns of basic frequencies by comprising a means of storing the time-varying patterns of basic frequencies for the fixed information extracted from natural voices upon employing a style of parameters of approximate models of the time-varying patterns of basic frequencies, a means of reading a parameter that suits a sentence to be inputted from the storing means, and a means of generating chronological orders of basic frequencies by using the parameter.

[Claim 4]

A voice synthesizing device, wherein the second generating means according to Claim 1 generates the time-varying patterns of basic frequencies by comprising a means of storing the time-varying pattern of basic extracted from natural voices for a combination of number of syllables and type of accent for the variable information upon employing a style of chronological orders of basic frequencies, and a means of selecting and reading a chronological order of basic frequency that suits a sentence

to be inputted from the storing means.

[Claim 5]

A voice synthesizing device, wherein the second generating means according to Claim 1 generates the time-varying patterns of basic frequencies by comprising a means of storing the time-varying patterns of basic frequencies extracted from natural voices for all combinations of number of syllables and type of accent for the variable information upon employing a style of parameters of approximate models of the time-varying patterns of basic frequencies, a means of selecting to read a parameter suitable to be inputted from the storing means, and a means of generating chronological orders of the basic frequency by using the parameter.

[Claim 6]

A voice synthesizing means device, wherein the second generating means according to Claim 1 includes a means of generating the time-varying patterns of basic frequencies for the variable information according to rules.

[Claim 7]

A voice synthesizing device for generating a duration time length which is a line of respective time lengths of synthesizing units, the voice synthesizing device comprising a first generating means of generating duration time lengths for fixed information, a second generating means of generating duration time lengths for variable information, an editing means of

generating a duration time length for a sentence upon sequentially connecting the duration time lengths generated by the respective generating means, and a means of synthesizing a voice signal by using the duration time length.

[Claim 8]

A voice synthesizing device, wherein the first generating means according to Claim 7 generates the duration time lengths by comprising a means of storing the duration time lengths for the fixed information extracted from natural voice and a means of reading a duration time length that suits a sentence to be inputted from the storing means.

[Claim 9]

A voice synthesizing device, wherein the second generating means according to Claim 7 includes a generating means of generating duration time lengths for variable information.

[Claim 10]

The voice synthesizing device according to Claim 1 or 7, wherein the voice synthesizing device comprises a text inputting means of enabling separation of fixed information and variable information wherein the voice synthesizing device presents fixed information while a synthesized sentence is inputted by a user with an user interface for inputting and editing variable information.

[Claim 11]

The voice synthesizing device according to Claim 1 or 7,

wherein the voice synthesizing device comprises a selecting means in which fixed information as well as input candidates for variable information are presented for designating variable information from among these candidates and a text inputting means of enabling separation of fixed information and variable information.

[Detailed Explanation of the Invention]

[0001]

[Industrially Applicable Field]

The present invention relates to a voice synthesizing device, and particularly to a voice synthesizing device for synthesizing voices used, for instance, for voice services such as traffic information or general weather condition, the voices being comprised of fixed information that are common to all groups of messages to be synthesized (hereinafter referred to as "fixed form parts") and variable information that are not common to the groups of messages (hereinafter referred to as "non-fixed form parts").

[0002]

Demands for reduction of labor and mechanization in general public became increasingly higher in these years. The field of various voice services is not an exception, and voice synthesizing devices are currently being used for voice services such as traffic information or general weather conditions or payment reference services in banks. It is therefore required

that such voice synthesizing devices provide synthesized voices with natural meters that are easy to hear.

[0003]

[Prior Art]

A conventional voice synthesizing device employs, for the fixed form part, a recording-editing method in which preliminarily recorded voices are reproduced or an analyzing-synthesizing method in which such voices converted into some kind of voice parameters are accumulated for synthesizing voices by using these parameters. For the non-fixed form parts represented by proper nouns or numerals, a ruled-synthesizing method was generally employed for synthesizing voices according to rules from character strings for connecting or switching voices that have been synthesized through the respective methods for output.

[0004]

A structural view of the voice synthesizing device according to the prior art is illustrated in Fig. 9. In the drawing, 1 denotes a text inputting means, 2 is a text analyzing means, 3 is a fixed form part synthesizing means, 4 is a non-fixed form part synthesizing means, 5 is an output voice connecting means, and 6 is a voice outputting means, respectively. A text inputted into the text inputting means 1 is analyzed in the text analyzing means 2 while referring to a word dictionary. Parts of the fixed form parts are consequently inputted into the fixed

form part synthesizing means 3 and voices are synthesized using accumulated voice data for the fixed form parts. Parts comprised of variable information are inputted into the non-fixed form part synthesizing means 4 to perform ruled-synthesizing from character strings. Voices that have been synthesized in the respective synthesizing means are connected in the output voice connecting means 5 such that these voices are connected as a sentence and are outputted through the voice outputting means 6.

[0005]

[Subject the Invention is to Solve]

However, in considering quality of voices, the quality of voices generated by the ruled-synthesizing method is inferior to those generated by the recording-editing method or the analyzing-synthesizing method in the present state.

[0006]

Thus, drawbacks were presented in that gaps in qualities of fixed form parts and non-fixed form parts were found in voices obtained by connecting fixed form parts obtained using the recording-editing method or the analyzing-synthesizing method with non-fixed form parts obtained using the ruled-synthesizing method and in that the non-fixed form parts including the important information within the sentences were difficult to be caught. On the contrary, it is easier to catch voices in which the whole sentence is generated to be of identical quality,

and thanks to improvements in qualities of voices of ruled-synthesizing method in these years accompanying technical improvements, sentences entirely synthesized through ruled-synthesizing method have become also sufficiently acceptable for actual use. Employing the ruled-synthesizing method for all, it will of course relieve the bother of rerecording voices in case the fixed form parts are to be changed.

[0007]

In case voices are to be synthesized from sentences in which Chinese characters and kana (Japanese syllabary) are mixed as we usually use in daily life, it is required to generate natural meters (intonation, accent, pause, etc.) while referring to the dictionary and rules when using the ruled-synthesizing method unlike the recording-editing method or the analyzing-synthesizing method. The following two problems are found in such a process.

[0008]

The first problem is found in a process of generating a phonetic character string by analyzing a sentence made up by mixing Chinese characters and kana. In this context, the term "phonetic character string" indicates a character string including notations indicative of positions of pauses or positions of accents in a phonemic string (substantially equal to "Romaji" (Latin alphabet) notation in the Japanese language) or in a syllabic string (substantially equal to kana-letter

notation in the Japanese language). Since the Japanese language is not written as being separated by each word and a single Chinese character may be read in many ways, erroneous reading, errors in accents or insertion of unnatural pauses may be happened frequently when trying to generate phonetic character strings while referring to dictionaries and rules. Although the first problem is solved by performing ruled-synthesizing upon character strings extracted from input string files for voice conversion serving as a storing means storing therein preliminarily generated input character strings including meter information (reference should be made to Japanese Patent Unexamined Publication No. 4-107598), it is required to reduce structural costs.

[0009]

The second problem occurs in a process of generating acoustic (physical) parameters from phonetic character strings. For instance, intonations, which are variations in heights of voices, are generally controlled by using time-varying patterns of basic frequencies having minimum frequencies included in voices of voiced sounds (hereinafter referred to as "F0 patterns"). This may be represented as chronological orders of basic frequencies of every several milliseconds (msec). Although Fujisaki models or dot-pitch models are well-known rules for generating such F0 patterns from the above phonetic character string, it is difficult to obtain such F0 patterns, which

delicately change depending on complicated human mechanisms of voice production, contents or meanings, on the basis of simple rules. Time lengths of respective phonemes or syllables are further set to suitable values such that generated voices will not clog or elongated but be natural. However, such time lengths cannot be ambiguously defined depending on types of phonemes or syllables but are complexly affected through positions of the phonemes or syllables within a sentence or through peripheral phonemic circumstances, and thus can not be obtained through simple rules, either.

[0010]

[Means of Solving the Subject]

Fig. 2 is a conceptual diagram of the present invention. Hereinafter, explanations will be made based on an exemplary sentence saying "TONIGHT'S WEATHER OF [TOKYO] DISTRICT WILL BE [FINE]".

[0011]

The sentence is comprised of a fixed form part "TONIGHT'S WEATHER OF DISTRICT ... WILL BE ..." and non-fixed form parts [TOKYO] and [FINE], wherein the non-fixed form parts may be respectively replaced with words such as [KANAGAWA] and [RAINY]. As for the fixed form part in synthesizing such a sentence, a F0 pattern and a duration time length for the fixed form part are extracted on the basis of a voice as uttered by a human reading the same sentence, which are stored as chronological orders of

basic frequency values of every several msec in case it is a F0 pattern or as a line of lengths of respective phonemes in case it is a duration time length. As for the non-fixed form part, F0 patterns for all combinations for the number of syllables of words or phrases that are expected to be inputted into the non-fixed form part as well as type of accent are stored, and a F0 pattern of a combination of the same number of syllables and type of accent is read on the basis of the inputted sentence or phonetic character strings obtained by analyzing the same. Since such F0 patterns are determined not only on the basis of the number of syllables and type of accent but also in view of a F0 pattern of the whole sentence, the F0 patterns will be respectively different and needs to be selected depending on the position within the fixed form part into which the same is inserted. For instance, the word "TOKYO" is of 4 molar 0 type, an F0 pattern of 4 molar 0 type is selected from among patterns to be inserted into the position of the fixed form part "TONIGHT'S ... DISTRICT". The duration time lengths for the non-fixed form parts are generated according to rules. By sequentially connecting F0 patterns and duration time lengths as retrieved (or generated) separately for the fixed form part and non-fixed form parts, a F0 pattern for the entire sentence is generated. The F0 patterns are connected in continuation within the entire sentences.

[0012]

Upon generating non-fixed form parts according to rules without storing F0 patterns thereof, it will be possible to obtain voices of even higher quality than that in cases all F0 patterns of entire sentences are generated according to rules.

[0013]

[Action]

A principle view of the present invention is shown in Fig. 1. In the drawing, 1 denotes a text input means, 7 denotes a text analyzing means, 8 denotes a means of generating F0 patterns/duration time lengths for the fixed form parts, 9 denotes a means of generating F0 patterns/duration time lengths for the non-fixed form parts, 10 denotes a means of connecting and editing F0 patterns/duration time lengths, 11 denotes an acoustic parameter generating means, 12 denotes a voice signal generating means, and 6 denotes a voice outputting means, respectively. A text to be synthesized is inputted into the text inputting means 1. The text analyzing means 7 separates the input text into non-fixed form parts and fixed form parts. While text analysis is necessary used for ruled-synthesizing of an arbitrary sentence for the separation into the non-fixed form parts and the fixed form parts in case the inputted text is an ordinary sentence in which Chinese characters and kana are mixed, if it is possible to separately input the non-fixed form parts and the fixed form parts through the user interface, the fixed form parts and the non-fixed form parts shall be simply

inputted into the respective means of generating F0 patterns/duration time lengths. The text analyzing means 7 further generates phonetic character strings (phonemic strings or syllabic strings) based on the inputted sentence to output to the acoustic parameter generating means 11. F0 patterns and duration time lengths are respectively generated for the fixed form parts and non-fixed form parts in the means of generating F0 patterns/duration time lengths for the fixed form parts 8 for the fixed form parts and in the means of generating F0 patterns/duration time lengths for the non-fixed form parts 9 for the non-fixed form parts. These F0 patterns and duration time lengths are sequentially connected in the means of connecting and editing F0 patterns/duration time lengths 10 to generate a F0 pattern and duration time length for the entire sentence. The acoustic parameter generating means 11 generates acoustic parameters such as formants on the basis of phonetic character strings such as phonemic strings or syllabic strings. The acoustic parameters are determined by the synthesizing method employed in the voice signal generating means 12. The synthesizing method may be a waveform editing method in which waveforms are directly edited, wherein wavelength connecting information are generated as equivalents instead of acoustic parameters in such a method, and such information are regarded herein to be included in acoustic parameters. The voice signal generating means 12 generates voice signals from the F0 patterns,

duration time lengths, and acoustic parameters to be outputted from the voice outputting means 6.

[0014]

[Embodiments]

It is considered that there are three levels for the F0 pattern generating method. The first level is a method in which F0 patterns extracted from natural voices are accumulated in a style of chronological orders of basic frequencies as they are which are then read at the time of synthesizing, and is a method with which it is expected to synthesize the most natural voices. The second level is a method in which F0 patterns of natural voices are approximated to models such that parameters of such models are accumulated, wherein these parameters are converted into the style of chronological orders of basic frequencies at the time of synthesizing. The third level is a method in which parameters of models are generated regularly on the basis of text analyzing results for generating chronological orders of basic frequencies from these parameters.

[0015]

It is further considered that there are two levels for the method for generating duration time lengths. The first level is a method in which duration time lengths extracted from natural voices are accumulated in lines of time lengths that are maintained as they are which are then read at the time of synthesizing. The second level is a method in which the time

lengths are generated regularly on the basis of text analyzing results. Various combinations of the above levels may be considered as methods for generating F0 patterns and duration time lengths for the non-fixed form parts and the fixed form parts. These will be described below as embodiments.

[0016]

A structural view of a first embodiment of the present invention is illustrated in Fig. 3. This embodiment corresponds to Claims 2, 4, 8 and 9. In the drawing, 011 denotes a text inputting unit, 71 denotes a text analyzing unit, 72 denotes a fixed form/non-fixed form determining unit, 73 denotes an output switching unit, 74 denotes a word dictionary, 75 denotes a unit for accumulating examples of sentences for the fixed form parts, 81 denotes a unit for reading duration time lengths for the fixed form parts, 82 denotes a unit for reading F0 patterns for the fixed form parts, 83 denotes a unit for accumulating duration time lengths for the fixed form parts, 84 denotes a unit for accumulating F0 patterns for the fixed form parts, 91 denotes a unit for generating duration time lengths for the non-fixed form parts, 92 denotes a unit for reading F0 patterns for the non-fixed form parts, 93 denotes an accent dictionary, 94 denotes a unit for accumulating F0 patterns for the non-fixed form parts, 101 denotes a unit for connecting and editing duration time lengths, 102 denotes a unit for connecting and editing F0 patterns, 111 denotes an acoustic parameter generating unit,

112 denotes an acoustic parameter accumulating unit, 121 denotes a voice signal generating unit and 61 denotes a voice outputting unit.

[0017]

F0 patterns for fixed form parts preliminarily extracted from natural voices for the fixed form parts are stored in the unit for accumulating F0 patterns for the fixed form parts 84, whereas for the non-fixed form parts, all combinations of the number of syllables and type of accent of F0 patterns for non-fixed form parts are stored in the unit for accumulating F0 patterns for the non-fixed form parts 94, and duration time lengths for fixed form parts extracted from natural voices are further stored in the unit for accumulating duration time lengths for the fixed form parts 83 for the fixed form parts. A text to be synthesized is inputted into the text inputting unit 011. In case a phonetic expression with Chinese characters and kana being mixed is inputted, the text is analyzed in the text analyzing part 71 while referring to the word dictionary 74. In the fixed form/non-fixed form determining unit 72, reference is made to examples of fixed form sentences stored in the unit for accumulating examples of sentences for the fixed form parts 75 so as to separate the result upon analysis into fixed form parts and non-fixed form parts. The output switching unit 73 outputs the fixed form parts and the non-fixed form parts to respective units for generating duration time lengths and F0 patterns. At

the same time, a phonetic character string of the input text (such as phonemic string or syllabic string) is outputted to the acoustic parameter generating part 111 as a result of analyzing the text.

[0018]

As for the fixed form parts, the unit for reading duration time lengths for the fixed form parts 81 reads duration time lengths from the unit for accumulating duration time lengths for the fixed form parts 83, while the unit for reading F0 patterns for the fixed form parts 82 reads F0 patterns from the unit for accumulating F0 patterns for the fixed form parts 84. They are respectively outputted to the unit for connecting and editing F0 patterns 102 upon passing the unit for connecting and editing duration time lengths 101. As for the non-fixed form parts, the unit for generating duration time lengths for the non-fixed form parts 91 generates duration time lengths according to rules. Generation of duration time lengths according to rules is generally performed by using a method in which a time length table is retrieved for each of the phonemes or syllables of the non-fixed form parts and are then corrected depending on phonemic circumstances or the like. Next, the unit for reading F0 patterns for the non-fixed form parts 92 acquires accents of the words of the non-fixed form parts from the accent dictionary 93, refers to the unit for accumulating F0 patterns for the non-fixed form parts 94 on the basis of the number of syllables and type of

accent and outputs the read F0 patterns to the unit for connecting and editing duration time lengths 101 and the unit for connecting and editing F0 patterns 102. The unit for connecting and editing duration time lengths 101 sequentially connects respective phonemic time lengths for the fixed form parts and non-fixed form parts so as to form a line for the duration time length of the entire sentence. The unit for connecting and editing F0 patterns 102 sequentially connects respective F0 patterns for the fixed form parts and non-fixed form parts so as to form a F0 pattern for the entire sentence. Since F0 patterns continue during utterance, in case non-succeeding portions shall be present in the respective F0 patterns as read for both, the fixed form parts and non-fixed form parts, editing such as suitable smoothing needs to be performed.

[0019]

On the other hand, the acoustic parameter generating part 111 generates acoustic parameters based on input phonetic character strings. The acoustic parameter accumulating part 112 stores acoustic parameters therein. The term "acoustic parameters" as used herein indicates voice data expressed by numeric values by using voice generating models in order to compress data capacity, and various types such as formants, PARCOR or LSP are known. Synthesizing methods using such acoustic parameters are respectively referred to as formant synthesis, PARCOR synthesis or LSP synthesis and are realized

by the voice signal generating unit 121. In addition, Another synthesizing method is a waveforms editing method in which waveforms are directly edited, though waveform connecting information are generated as equivalents instead of acoustic parameters in such a method, such information are regarded herein to be included in acoustic parameters. Acoustic parameters are accumulated by phonetic characters or such more diversified units depending on front and rear phonemic circumstances. By reading and connecting these in accordance with phonetic character strings, acoustic parameter strings of the synthesized sentence can be generated. The voice signal generating unit 121 generates voice signals using the duration time lengths, F0 patterns and acoustic parameter strings for the synthesized sentence generated above. The voice outputting unit 61 outputs the voice signal as a synthesized voice upon performing DA conversion.

[0020]

A structural view of the second embodiment of the present invention is illustrated in Fig. 4. This embodiment corresponds to Claims 3 and 5. The present embodiment is arranged in that the unit for reading F0 patterns for the fixed form parts 82 and the unit for accumulating F0 patterns for the fixed form parts 84 are substituted by a unit for reading F0 parameters for the fixed form parts 85 and a unit for generating F0 patterns for the fixed form parts 86, and a unit for accumulating F0 parameters for the fixed form parts 87, while the unit for reading

F0 patterns for the non-fixed form parts 92 and the unit for accumulating F0 patterns for the non-fixed form parts 94 are substituted by a unit for reading F0 parameters for the non-fixed form parts 95 and a unit for generating F0 patterns for the non-fixed form parts 96, and a unit for accumulating F0 parameters for the non-fixed form parts 97.

[0021]

In this embodiment, F0 patterns that have been extracted preliminarily from natural voices are approximated through models so that the parameters may be accumulated in the unit for accumulating F0 parameters for the fixed form parts 87 and the unit for accumulating F0 parameters for the non-fixed form parts 97. In synthesizing voices, for the fixed form parts, the unit for reading F0 parameters for the fixed form parts 85 reads F0 patterns for the fixed form parts from the unit for accumulating F0 parameters for the fixed form parts 87, and the unit for generating F0 patterns for the fixed form parts 86 generates chronological orders of basic frequencies (F0 patterns) from the parameters. Similarly, for the non-fixed form parts, the unit for reading F0 parameters for the non-fixed form parts 95 acquires accents of words of the non-fixed form parts from the accent dictionary 93 and reads proper F0 parameters from the unit for accumulating F0 parameters for the non-fixed form parts 97 depending on the number of syllables and type of accent, and the unit for generating F0 patterns for the non-fixed

form parts 96 generates chronological orders of basic frequencies (F0 patterns) from these.

[0022]

A structural view of the third embodiment of the present invention is illustrated in Fig. 5. This embodiment corresponds to Claim 6. The present embodiment is arranged in that the unit for reading F0 patterns for the non-fixed form parts 92 and the unit for accumulating F0 patterns for the non-fixed form parts 94 are substituted by a unit for generating F0 patterns for the non-fixed form parts 98. Since processes of the remaining units are identical to those of the first embodiment, only the unit for generating F0 patterns for the non-fixed form parts 98 will be explained here.

[0023]

The unit for generating F0 patterns for the non-fixed form parts 98 acquires accents of words of the non-fixed form parts from the accent dictionary 93 to generate F0 patterns according to rules with regard to their positions within a sentence. As a method for generating F0 patterns according to rules, a method employing models such as Fujisaki models or dot-pitch models is commonly used and these are applicable also in the present case.

[0024]

A structural view of the fourth embodiment of the present invention is illustrated in Fig. 6. This embodiment corresponds

to Claims 10 and 11. The present embodiment is arranged in that more accurate text analysis is enabled by substituting the user interface of the text inputting unit of the first embodiment. An input interface unit 012 reads fixed form parts from a unit for accumulating examples of sentences for the fixed form parts 013 and displays these as illustrated in Fig. 7 or 8 as a user interface. In Fig. 7, for the fixed form parts are provided with columns having display functions only while the non-fixed form parts are provided with columns having editing functions capable of inputting/editing words freely and makes a user to input non-fixed form parts. By performing input using such an interface, it will not be required for performing determination of fixed form parts and non-fixed form parts and text analysis may be performed by retrieving only fixed form parts using the word dictionary 74.

[0025]

In Fig. 8, the device includes an interface in which input candidates for the non-fixed form parts are accumulated in the unit for accumulating examples of sentences for the fixed form parts 13, and upon designating a column of a non-fixed form part, input candidates to be inputted into this spot are displayed so that it is possible to designate which of them is inputted using a candidate selecting means. Similarly, determination of fixed form parts and non-fixed form parts is not necessary and text analysis may be performed by retrieving only fixed form

parts using the word dictionary 74. Sequent processes are identical to those of the other embodiment.

[0026]

[Effect of the Invention]

As explained above, according to the present invention, it is possible to synthesize voices with natural meters that are easy to hear in a voice synthesizing device for synthesizing voices of fixed sentences which is particularly used for voice services such as traffic information or general weather condition.

[Brief Explanation of the Drawings]

[Fig. 1]

It shows a principle view of the present invention.

[Fig. 2]

It shows a conceptual view illustrating basic ideas of the present invention.

[Fig. 3]

It shows a first embodiment of the present invention.

[Fig. 4]

It shows a second embodiment of the present invention.

[Fig. 5]

It shows a third embodiment of the present invention.

[Fig. 6]

It shows a fourth embodiment of the present invention.

[Fig. 7]

It shows a first example of a user interface of the present invention.

[Fig. 8]

It shows a second example of a user interface of the present invention.

[Fig. 9]

It shows a prior art.

[Explanations of Reference Numerals]

1 Text inputting means

2, 7 Text analyzing means

3 Fixed form part synthesizing means

4 Non-fixed form part synthesizing means

5 Output voice connecting means

6 Voice outputting means

8 Means of generating F0 patterns/duration time lengths for the fixed form parts

9 Means of generating F0 patterns/duration time lengths for the non-fixed form parts

10 Means of connecting and editing F0 patterns/duration time lengths (abbreviated as editing means)

11 Acoustic parameter generating means

12 Voice signal generating means

61 Voice outputting unit

71, 71' Text analyzing unit

72 Fixed form/non-fixed form determining unit

73 Output switching unit

74 Word dictionary

75, 013 Unit for accumulating examples of sentences for the fixed form parts

81 Unit for reading duration time lengths for the fixed form parts

82 Unit for reading F0 patterns for the fixed form parts

83 Unit for accumulating duration time lengths for the fixed form parts

84 Unit for accumulating F0 patterns for the fixed form parts

85 Unit for reading F0 parameters for the fixed form parts

86 Unit for generating F0 patterns for the fixed form parts

87 Unit for accumulating F0 parameters for the fixed form parts

91 Unit for generating duration time lengths for the non-fixed form parts

92 Unit for reading F0 patterns for the non-fixed form parts

93 Accent dictionary

94 Unit for accumulating F0 patterns for the non-fixed form parts

95 Unit for reading F0 parameters for the non-fixed form parts

96, 98 Unit for generating F0 patterns for the non-fixed form parts

97 Unit for accumulating F0 parameters for the non-fixed form parts

011 Text inputting unit

012 Input interface unit

101 Unit for connecting and editing duration time lengths

102 Unit for connecting and editing F0 patterns

111 Acoustic parameter generating unit

112 Acoustic parameter accumulating unit

121 Voice signal generating unit

Translation of drawings

[FIG. 1]

(1) Principle view of the present invention

(2) Text inputting means

(3) Text analyzing means

(4) Means of generating F0 patterns/duration time lengths for the fixed form parts

(5) Means of generating F0 patterns/duration time lengths for the non-fixed form parts

(6) Means of connecting and editing F0 patterns/duration time lengths

(7) Acoustic parameter generating means

(8) Voice signal generating means

(9) Voice outputting means

[FIG. 2]

(10) Conceptual view illustrating basic ideas of the present invention

(11) Frequency

(12) Fixed form parts

(13) TONIGHT'S

(14) WEATHER OF ... DISTRICT

(15) WILL BE

(16) Time

(17) Non-fixed form parts

(18) 4 molar 0 type pattern

(19) 2 molar 1 type pattern

(20) TOKYO

(21) KANAGAWA PREFECTURE

(22) FINE

(23) RAINY

[FIG. 3]

(1) First embodiment of the present invention

(2) Text inputting unit

(3) Text analyzing unit

(4) Fixed form/non-fixed form determining unit

(5) Output switching unit

(6) Word dictionary

(7) Unit for accumulating examples of sentences for the fixed form parts

(8) Unit for accumulating duration time lengths for the fixed form parts

(9) Unit for accumulating F0 patterns for the fixed form parts

(10) Unit for reading duration time lengths for the fixed form parts

(11) Unit for reading F0 patterns for the fixed form parts

(12) Unit for generating duration time lengths for the non-fixed form parts

(13) Unit for reading F0 patterns for the non-fixed form parts

(14) Accent dictionary

(15) Unit for accumulating F0 patterns for the non-fixed form

parts

(16) Unit for connecting and editing duration time lengths

(17) Unit for connecting and editing F0 patterns

(18) Acoustic parameter generating unit

(19) Acoustic parameter accumulating unit

(20) Voice signal generating unit

(21) Voice outputting unit

[FIG. 4]

(1) Second Embodiment of the present invention

(2) From output switching unit 73

(3) Unit for accumulating duration time lengths for the fixed form parts

(4) Unit for accumulating F0 parameters for the fixed form parts

(5) Unit for reading duration time lengths for the fixed form parts

(6) Unit for reading F0 parameters for the fixed form parts

(7) Unit for generating F0 patterns for the fixed form parts

(8) Unit for generating duration time lengths for the non-fixed form parts

(9) Unit for reading F0 parameters for the non-fixed form parts

(10) Unit for generating F0 patterns for the non-fixed form parts

(11) Accent dictionary

(12) Unit for accumulating F0 parameters for the non-fixed form

parts

(13) To unit for connecting and editing duration time lengths 101

[FIG. 5]

(14) Third Embodiment of the present invention

(15) From output switching unit 73

(16) Unit for generating duration time lengths for the non-fixed form parts

(17) Unit for generating F0 patterns for the non-fixed form parts

(18) Accent dictionary

(19) To unit for connecting and editing duration time lengths 101

[FIG. 6]

(20) Fourth embodiment of the present invention

(21) Input interface unit

(22) Unit for accumulating examples of sentences for the fixed form parts

(23) Text analyzing unit

(24) Output switching unit

(25) Word dictionary

(26) To unit for reading duration time lengths for the fixed form parts 81

(27) To unit for reading duration time lengths for the non-fixed form parts 91

(28) To acoustic parameter generating unit 111

[FIG. 7]

(29) First example of a user interface of the present invention

(30) Non-fixed form parts (editable part)

(31) Fixed form parts

(32) TONIGHT'S WEATHER OF [TOKYO] DISTRICT WILL BE [FINE].

[FIG. 9]

(33) Prior Art

(34) Text inputting means

(35) Text analyzing means

(36) Fixed form part synthesizing means

(37) Non-fixed form part synthesizing means

(38) Output voice connecting means

(39) Voice outputting means

[FIG. 8]

(40) Second example of user interface of the present invention

(41) TONIGHT'S WEATHER OF [...] DISTRICT WILL BE [FINE]

(42) TOKYO/KANAGAWA/SAITAMA/SOUTHERN CHIBA/NORTHERN CHIBA/SOUTHERN IBARAGI/NORTHERN IBARAGI/SOUTHERN TOCHIGI/NORTHERN TOCHIGI/OTHERS

る。以降の処理は他の実施例と同様である。

【００２６】

【発明の効果】以上説明した様に、本発明によれば、交通情報や天気概況の音声サービスなどに用いる、定型文音声を合成するための音声合成装置において、聞き取りやすく、自然な韻律をもつ音声を合成することができる。

【図面の簡単な説明】

【図１】　本発明の原理図である。

【図２】　本発明の基本的な考え方を示した概念図である。

【図３】　本発明の第１の実施例である。

【図４】　本発明の第２の実施例である。

【図５】　本発明の第３の実施例である。

【図６】　本発明の第４の実施例である。

【図７】　本発明のユーザインターフェースの第１の例である。

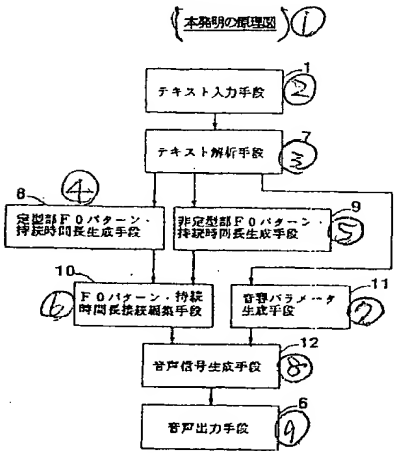
【図８】　本発明のユーザインターフェースの第２の例である。

【図９】　従来例である。

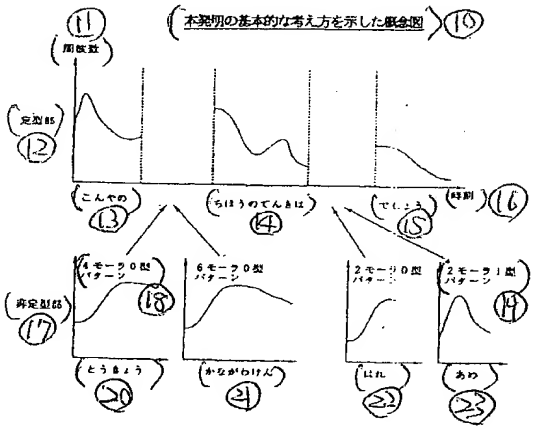
【符号の説明】

１　テキスト入力手段
２、７　テキスト解析手段
３　定型部合成手段
４　非定型部合成手段
５　出力音声接続手段
６　音声出力手段
８　定型部Ｆ０パターン・持続時間長生成手段
９　非定型部Ｆ０パターン・持続時間長生成手段
１０　Ｆ０パターン・持続時間長接続編集手段（編集手＊段と略す。）
１１　音響パラメータ生成手段
１２　音声信号生成手段
６１　音声出力部
７１、７１'　テキスト解析部
７２　定型／非定型判定部
７３　出力切替部
７４　単語辞書
７５、０１３　定型部文例蓄積部
８１　定型部持続時間長読み込み部
８２　定型部Ｆ０パターン読み込み部
８３　定型部持続時間長蓄積部
８４　定型部Ｆ０パターン蓄積部
８５　定型部Ｆ０パラメータ読み込み部
８６　定型部Ｆ０パターン生成部
８７　定型部Ｆ０パラメータ蓄積部
９１　非定型部持続時間長生成部
９２　非定型部Ｆ０パターン読み込み部
９３　アクセント辞書
９４　非定型部Ｆ０パターン蓄積部
９５　非定型部Ｆ０パラメータ読み込み部
９６、９８　非定型部Ｆ０パターン生成部
９７　非定型部Ｆ０パラメータ蓄積部
０１１　テキスト入力部
０１２　入力インターフェース部
１０１　持続時間長接続編集部
１０２　Ｆ０パターン接続編集部
１１１　音響パラメータ生成部
１１２　音響パラメータ蓄積部
１２１　音声信号生成部

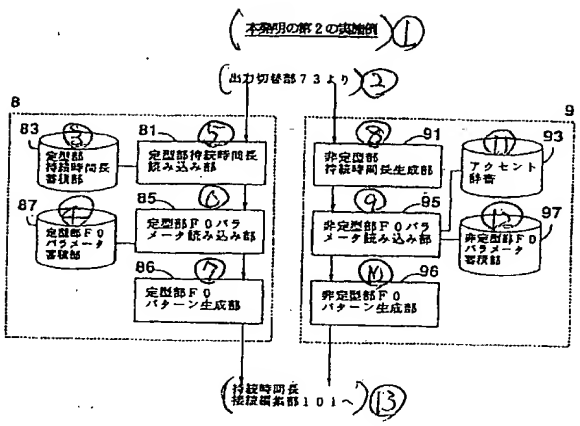
【図１】



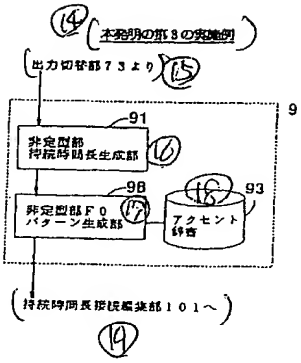
【図２】



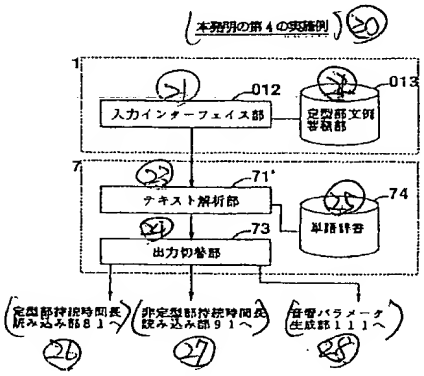
【図４】


本発明の第２の実施例
出力切替部７３より
持続時間長接続編集部１０１へ
定型部持続時間長読み込み部
定型部ＦＯパラメータ読み込み部
定型部ＦＯパターン生成部
非定型部持続時間長生成部
非定型部ＦＯパラメータ読み込み部
非定型部ＦＯパターン生成部
アクセント辞書
非定型部ＦＯパラメータ蓄積部

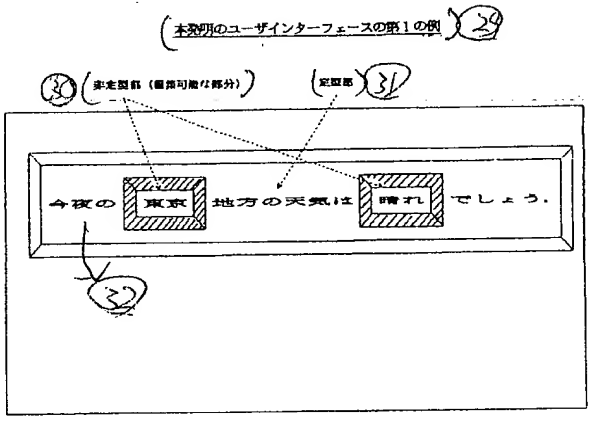
【図５】


本発明の第３の実施例
出力切替部７３より
非定型部持続時間長生成部
非定型部ＦＯパターン生成部
アクセント辞書
持続時間長接続編集部１０１へ

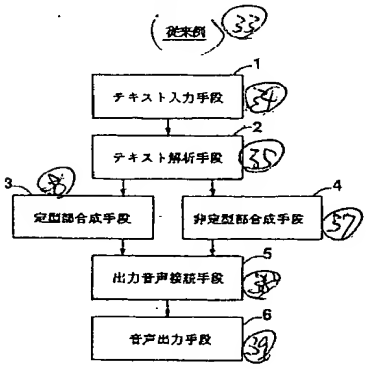
【図６】


本発明の第４の実施例
入力インターフェイス部
定型部文例蓄積部
テキスト解析部
単語辞書
出力切替部
定型部持続時間長読み込み部８１へ
非定型部持続時間長読み込み部９１へ
音響パラメータ生成部１１１へ

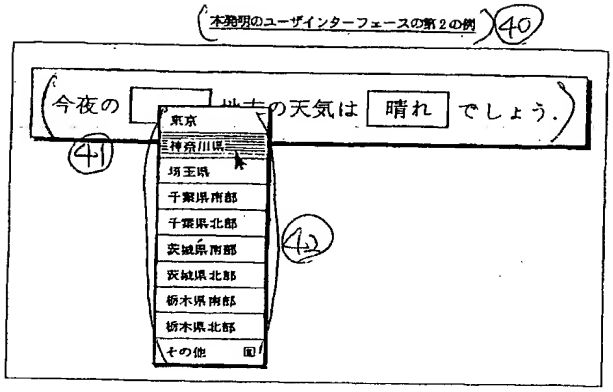
【図７】


本発明のユーザインターフェースの第１の例
非定型部（編集可能な部分）
定型部
今夜の
東京
地方の天気は
晴れ
でしょう．

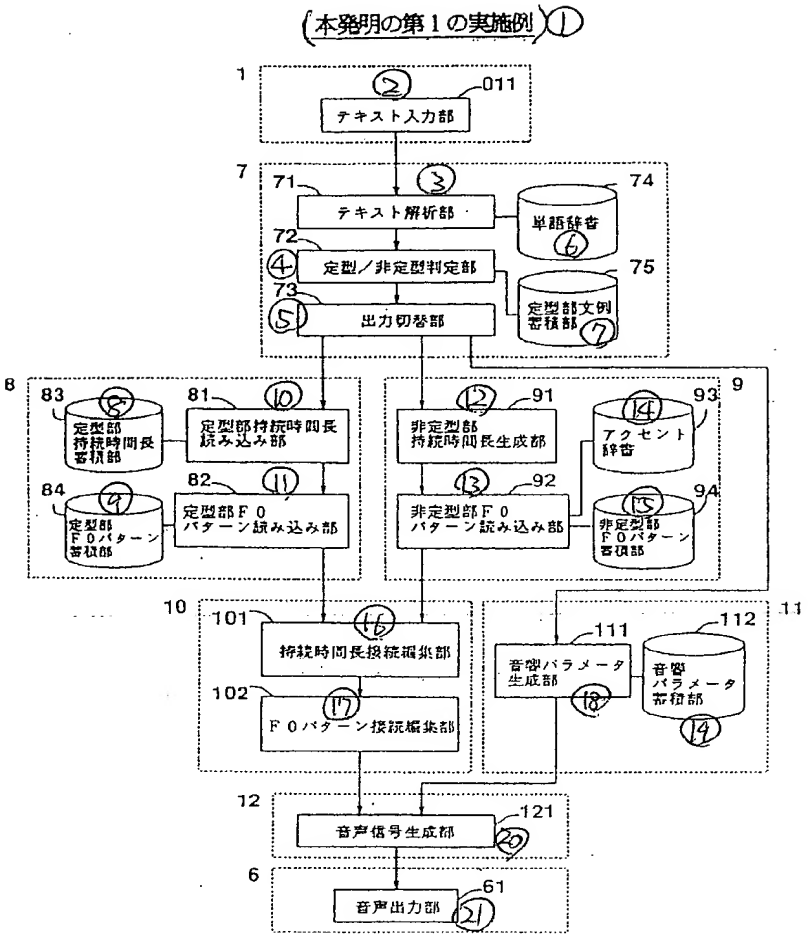
【図９】


従来例
テキスト入力手段
テキスト解析手段
定型部合成手段
非定型部合成手段
出力音声接続手段
音声出力手段

【図８】


本発明のユーザインターフェースの第２の例
40
今夜の
地方の天気は
晴れ
でしょう.
41
東京
神奈川県
埼玉県
千葉県南部
千葉県北部
茨城県南部
茨城県北部
栃木県南部
栃木県北部
その他
42

〔図３〕


本発明の第１の実施例 ①
1
②
011
テキスト入力部
7
71
③
テキスト解析部
74
単語辞書 ⑥
72
④
定型／非定型判定部
75
73
⑤
出力切替部
定型部文例蓄積部 ⑦
8
83
⑧
定型部持続時間長蓄積部
81
⑩
定型部持続時間長読み込み部
84
⑨
定型部Ｆ０パターン蓄積部
82
⑪
定型部Ｆ０パターン読み込み部
9
⑫
91
非定型部持続時間長生成部
⑭
93
アクセント辞書
⑬
92
非定型部Ｆ０パターン読み込み部
⑮
94
非定型部Ｆ０パターン蓄積部
10
101
⑯
持続時間長接続編集部
102
⑰
Ｆ０パターン接続編集部
11
111
音韻パラメータ生成部 ⑱
112
音韻パラメータ蓄積部 ⑭
12
121
音声信号生成部 ⑳
6
61
音声出力部 ㉑